

WHAT IS CLAIMED IS:

1. A system for analyzing a bio chip comprising :
a GO(gene ontology) term assigning part for receiving a statistical clustering data
5 obtained from the empirical results of the bio chip, and assigning relevant GO terms
to every gene contained in each cluster;
a GO code converting part for converting the GO terms assigned by the GO term
assigning part to the genes into GO codes, the GO code comprising a group of
predetermined numbers; and
10 a biological meaning extracting part for calculating pseudo distances between one
of GO terms contained in a predetermined group on GO tree structure and the GO
terms corresponding to the genes contained in the cluster, and calculating at least one
of average pseudo distance or maximum pseudo distance of the calculated pseudo
distances, and calculating at least one of average pseudo distances or maximum
15 pseudo distances for all GO terms included in the predetermined group on GO tree
structure and the GO terms corresponding to the genes contained in the cluster, and
determining an optimum GO term matching with the cluster.
2. The system according to claim 1, wherein the GO term assigning part assigns
20 GO terms to the genes using biology database mining.
3. The system according to claim 1, wherein the GO code converting part converts
the GO terms into the GO codes according to a level of a GO term, a parent-node of
the GO term and an order of the GO term in the level.
25
4. The system according to claim 1, wherein the biological meaning extracting
part comprises :
an optimum cross-point extracting part for extracting optimum cross-points
between the GO terms on the GO tree structure and the GO terms assigned to the
30 genes contained in the predetermined group;
a pseudo distance calculating part for calculating pseudo distances between the

GO terms on the GO tree structure and the GO terms assigned to the genes contained in the cluster by using the optimum cross-points information;

an average pseudo distance calculating part for calculating average pseudo distance of the pseudo distances calculated from the pseudo distance calculating part;

5 a maximum pseudo distance determining part for determining maximum distance among the pseudo distances calculated from the pseudo distance calculating part; and

an optimum matching node determining part for comparing average pseudo distances or maximum pseudo distances for all GO terms contained in the predetermined group, and determining a GO term with minimum value of the
10 average pseudo distance or of the maximum pseudo distance to be optimum matching node of the cluster.

5. The system according to claim 4, wherein the GO terms contained in the predetermined group are all terms on the GO tree structure.

15

6. The system according to claim 4, wherein the GO terms contained in the predetermined group are GO terms included in a selected level on the GO tree structure.

20 7. The system according to claim 4, wherein the optimum cross-point extracting part determines a GO term in the lowest level among GO terms which include two GO terms in a lower level on the GO tree structure to be the optimum cross-point.

8. The system according to claim 1, wherein the GO tree structure comprises a level which a predetermined weight is granted to, and wherein the pseudo distance
25 calculated by the pseudo distance calculating part is the weight granted to a level where the optimum cross-point exists.

9. A method for analyzing a bio chip comprising :

a) receiving a statistical clustering data obtained from empirical results of the bio
30 chip to assign relevant GO terms to every gene contained in each cluster;

b) converting the GO terms assigned to the genes into GO codes, the GO code

comprising a group of predetermined numbers;

c) calculating pseudo distances between one of GO terms contained in a predetermined group on GO tree structure and the GO terms corresponding to the genes contained in the cluster by using the GO codes;

5 d) calculating at least one of average pseudo distance or maximum pseudo distance of the pseudo distances calculated in the step (c); and

e) repeating the step (c) and the step (d) for every GO term on the GO tree structure contained in the predetermined group to determine an optimum GO term matching with the cluster.

10

10. The method according to claim 9, wherein the step (a) assigns GO terms to the genes using biology databases mining.

11. The method according to claim 9, wherein the step (b) converts the GO terms
15 into the GO codes according to a level of a GO term, a parent-node of the GO term and an order of the GO term in the level.

12. The method according to claim 9, wherein the GO terms contained in the predetermined group are all terms on the GO tree structure.

20 13. The method according to claim 9, wherein the GO terms contained in the predetermined group are GO terms included in a selected level on GO tree structure.

14. The method according to claim 9, wherein the step (c) comprises steps of:
extracting optimum cross-points between the GO terms on the GO tree structure
25 and the GO terms assigned to the genes contained in the cluster; and
calculating pseudo distances between the GO terms on the GO tree structure and the GO terms assigned to the genes contained in the cluster by using the optimum cross-points information.

30 15. The method according to claim 9, wherein the step (e) determines a GO term on the GO tree structure with minimum value of the average pseudo distance or the

maximum pseudo distance to be an optimum matching node of the cluster

16. The method according to claim 14, wherein the step for extracting the optimum cross-points determines a GO term in the lowest level among GO terms
5 which include two GO terms in lower level on the GO tree structure to be the optimum cross-point.

17. The method according to claim 14, wherein the GO tree structure comprises a level which a predetermined weight is granted to, and wherein the calculated pseudo
10 distance is an weight granted to a level where the optimum cross-point exists.

18. A digital device readable medium containing program instructions for executing an analysis of a bio chip, the medium comprising the program instructions for :

- 15 a) receiving a statistical clustering data obtained from empirical results of the bio chip, and for assigning relevant GO terms to every gene contained in each cluster;
- b) converting the GO terms assigned to the genes into GO codes, the GO code comprising a group of predetermined numbers;
- c) calculating pseudo distances between one of GO terms on GO tree structure
20 contained a predetermined group and the GO terms corresponding to the genes contained in the cluster by using the GO codes;
- d) calculating at least one of average pseudo distance or maximum pseudo distance of the pseudo distances calculated in the step (c); and
- e) repeating the step (c) and the step (d) for every GO term on the GO tree
25 structure contained in the predetermined group to determine an optimum GO term matching with the cluster.